Correspondence

Selection on Alu sequences?

John F.Y. Brookfield

The draft of the human genome project reveals an unexpected distribution of Alu interspersed repetitive sequences [1]. This has been interpreted as evidence that Alu sequences 'may benefit their human hosts' [1] and that they 'have a positive function' [2]. The implication is that the majority of Alu sequences increase the Darwinian fitness of their bearers. Here I suggest that this conclusion is inconsistent with our knowledge of human population genetics.

There is a relationship between the time since an Alu sequence was inserted into the genome and the GC content of its surrounding DNA. Alu sequences inferred to have inserted within the last five million years are slightly more abundant in low-GC, gene-poor regions, whereas older Alus, in classes inserted from 5 to 100 million years ago, are increasingly found in high-GC gene-rich regions. Contrary to the conclusions drawn in the report [1] and the associated News and Views article [2], this observation does not indicate that Alu sequences are advantageous to their human hosts. While one could imagine that Alu sequences active 50 million years ago were targeting GC-rich regions while today's Alus target AT-rich regions, a more parsimonious interpretation is to assume that the sequence's insertion preferences have been constant, and that the change in the sequence's relative abundance reflects a process in which, following insertion, it increases its relative abundance in GC-rich DNA with time. The most likely such process, and one considered and dismissed by the authors [1], is that deletions removing

Alu sequences from GC-rich DNA are likely to be harmful and prevented from spreading in the population by natural selection. This implies no functional importance for an Alu sequence itself, but merely that, as the deletions of Alus are very unlikely to be precise, a deletion event removing an Alu is also likely to remove valuable sequences around it, and the chromosome bearing the deletion will be lost by selection.

The explanation favoured by the authors for Alu enrichment in gene-rich regions is that of positive selection in favour of Alus in GC-rich DNA. This theory, however, cannot explain the observations. The data show that Alu sequences up to five million years old are not enriched in GC-rich regions. But in human population genetics, estimated times to common ancestry of typical genomic regions show that Alu sequences which are five million years old have already been fixed (found in all individuals) in the population. This observation is also what would be expected from neutrality and genetic drift, given the human effective population size. (Alu sequences which are truly advantageous will spread to fixation much more quickly.) Earlier human ancestors would also be expected to have had similar fixation times for Alu insertions. Yet it is only during the spread to fixation of Alu sequences that positive natural selection has any opportunity to act. Thus, an increasing abundance of Alu sequences in GC-rich DNA as they age beyond five million years cannot be the result of natural selection for positive functions of Alu insertions.

References

1. International Human Genome Sequencing Consortium: Initial sequencing and analysis of the human genome. *Nature* 2001, 409:860-921.
2. Baltimore D: Our genome unveiled. *Nature* 2001, 409:814-816.

Address: Institute of Genetics, University of Nottingham, Queens Medical Centre, Nottingham NG7 2UH, UK.
E-mail john.brookfield@nottingham.ac.uk